Space Alternating Variational Bayesian Learning for LMMSE Filtering

Christo Kurisummoottil Thomas Dirk Slock
EURECOM, Sophia-Antipolis, France, Email:{kurisumm,slock}@eurecom.fr

***Abstract*—In this paper, we address the fundamental problem of sparse signal recovery for temporally correlated multiple measurement vectors (MMV) in a Bayesian framework. The temporal correlation of the sparse vector is modeled using a first order autoregressive process. In the case of time varying sparse signals, conventional tracking methods like Kalman filtering fail to exploit the sparsity of the underlying signal. Moreover, the computational complexity associated with sparse Bayesian learning (SBL) renders it infeasible even for moderately large datasets. To address this issue, we utilize variational approximation technique (which allows to obtain analytical approximations to the posterior distributions of interest even when exact inference of these distributions is intractable) to propose a novel fast algorithm called space alternating variational estimation with Kalman filtering (SAVE-KF). Similarly as for SAGE (space-alternating generalized expectation maximization) compared to EM, the component-wise approach of VB appears to allow to avoid a lot of bad local optima, explaining the better performance, apart from lower complexity. Simulation results also show that the proposed algorithm has a faster convergence rate and achieves lower mean square error (MSE) than other state of the art fast SBL methods for temporally correlated measurement vectors.**

***Keywords*—** Sparse Bayesian Learning, Variational Bayes, Kalman Filtering

I. Introduction

Sparse signal reconstruction and compressed sensing (CS) has received significant attraction in the recent years. The compressed sensing problem can be formulated as,

$$\mathbf{y}_t = \mathbf{A}_t \mathbf{x}_t + \mathbf{w}_t, \tag{1}$$

where $\mathbf{y}_t$ is the observations or data at time t, $\mathbf{A}_t$ is called the measurement or the sensing matrix which is known and is of dimension $N \times M$ with $N < M$, $\mathbf{x}_t$ is the M-dimensional sparse signal and $\mathbf{w}_t$ is the additive noise. $\mathbf{x}_t$ contains only K non-zero entries, with $K << M$ and is modeled by an AR(1) (auto-regressive) process. $\mathbf{w}_t$ is assumed to be a white gaussian noise, $\mathbf{w}_t \sim \mathcal{N}(0, \gamma^{-1}\mathbf{I})$. Most of the MMV SBL algorithms are obtained by straightforward extension of the algorithms in the single measurement vector case (SMV). In SMV, to address this problem, a variety of algorithms such as the orthogonal matching pursuit [1], the basis pursuit method [2] and the iterative re-weighted l_1 and l_2 algorithms [3] exist in the literature. In a Bayesian setting, the aim is to calculate the posterior distribution of the parameters given some observations (data) and some a priori knowledge. Sparse Bayesian learning algorithm was first introduced by [4] and then proposed for the first time for sparse signal recovery by [5]. Performance can be further improved by exploiting the temporal correlation across the sparse vectors [6], [7]. However, most of these algorithms do offline or batch processing. Nevertheless the complexity of these solutions doesn't scale with the problem size. In order to render low complexity or low latency solutions, online processing algorithms (which processes small set of measurement vectors at any time) will be necessary.

In conventional CS, the time invariant sparse signal is estimated using less measurements than the size of the signal. On the other hand, in sparse adaptive estimation, a time varying signal $\mathbf{x}_t$ is estimated time-recursively by exploiting the sparsity property of the signal. Conventional adaptive filtering methods such as LMS or recursive least squares (RLS) doesn't exploit the underlying sparseness in the signal $\mathbf{x}_t$ to improve the estimation performance. Kalman filter focus on estimation of the dynamical state from noisy observations where the dynamic and measurement process are considered to be from linear Gaussian state space model. However, classical Kalman filtering assume complete knowledge of the apriori information about the model parameters and noise statistics.

In sparse Bayesian learning, the sparse signal $\mathbf{x}_t$ is modeled using a prior distribution $p(\mathbf{x}_t/\boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = [\alpha_1 \ ... \ \alpha_M]^T$ is the vector parameter and α_i is the inverse of the variance of $x_{t,i}$, also interpreted as the precision variable. Since most of the elements of $\mathbf{x}_t$ are zero, most of the α_i should be very high favoring solutions with few non-zero components.

In Type II maximum likelihood method or evidence procedure [8], an estimate of the parameters $\boldsymbol{\alpha}, \gamma$ and sparse signal $\mathbf{x}_t$ is done iteratively using evidence maximization. In [9], the authors propose a Fast Marginalized Maximum Likelihood (FMML) by alternating maximization of the hyperparameters α_i.

SBL involves a matrix inversion step at each iteration, which makes it a computationally complex algorithm even for moderately large datasets. An alternative approach to SBL is using variational approximation for Bayesian inference [10]–[13]. Variational Bayesian (VB) inference tries to find an approximation of the posterior distribution which maximizes the variational lower bound on $log\, p(\mathbf{y}_t)$. We propose a space alternating variational estimation based technique for single measurement vectors in [14].

A. Contributions of this paper

In this paper:

ISBN 978-90-827970-1-5 © EURASIP 2018

- We propose a novel Space Alternating Variational Estimation (SAVE) based SBL technique for LMMSE filtering called SAVE-KF. The proposed solution is for a multiple measurement case with an AR(1) process for the temporal correlation of the sparse signal. The update and prediction stages of the proposed algorithm reveals links to the Kalman filtering.
- Numerical results suggest that our proposed solution has a faster convergence rate (and hence lower complexity) than (even) the existing fast SBL and performs better than the existing fast SBL algorithms in terms of reconstruction error in the presence of noise.

In the following, boldface lower-case and upper-case characters denote vectors and matrices respectively. the operators $tr(\cdot)$, $(\cdot)^T$ represents trace,and transpose respectively. The operator $(\cdot)^H$ represents the conjugate transpose or conjugate for a matrix or a scalar respectively. A complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Theta}$ is distributed as $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Theta})$. $diag(\cdot)$ represents the diagonal matrix created by elements of a row or column vector. The operator $< x >$ or $\mathrm{E}(\cdot)$ represents the expectation of x. $||\cdot||$ represents the Frobenius norm. $\angle(\cdot)$ represents the angle or phase of a complex number. $\Re\{(\cdot)\}$ represents the real part of $(\cdot)$. All the variables are complex here unless specified otherwise.

II. State Space Model

Sparse signal $\mathbf{x}_t$ is modeled using an AR(1) process with correlation coefficient matrix $\mathbf{F}$, with $\mathbf{F}$ diagonal. The state space model can be written as follows,

$$\begin{aligned} \mathbf{x}_t &= \mathbf{F}\mathbf{x}_{t-1} + \mathbf{v}_t, \quad \text{State Update,} \\ \mathbf{y}_t &= \mathbf{A}_t\mathbf{x}_t + \mathbf{w}_t, \qquad \text{Observation,} \end{aligned} \tag{2}$$

where $\mathbf{x}_t = [x_{t,1}, \ ..., \ x_{t,M}]^T$. Matrices $\mathbf{F}$ and $\boldsymbol{\Gamma}$ are defined as,

$$\mathbf{F} = \begin{bmatrix} f_1 & 0 & \dots & 0 \\ 0 & f_2 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f_M \end{bmatrix}, \boldsymbol{\Gamma} = \begin{bmatrix} \frac{1}{\sqrt{\alpha_1}} & \dots & 0 \\ 0 & & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sqrt{\alpha_M}} \end{bmatrix}, \tag{3}$$

Here a_i represents the correlation coefficient and α_i represents the inverse variance of $x_{t,i} \sim \mathcal{CN}(0, \frac{1}{\alpha_i})$. Further, $\mathbf{v}_t \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Gamma}(\mathbf{I}-\mathbf{F}\mathbf{F}^H))$ and $\mathbf{w}_t \sim \mathcal{CN}(\mathbf{0}, \frac{1}{\gamma}\mathbf{I})$. $\mathbf{v}_t$ are the complex Gaussian mutually uncorrelated innovation sequences. $\mathbf{w}_t$ is independent of the innovation process $\mathbf{v}_t$. Further we define, $\boldsymbol{\Lambda} = \boldsymbol{\Gamma}(\mathbf{I} - \mathbf{F}\mathbf{F}^H) = \text{ diag } (\lambda_1, ..., \lambda_M)$.

Although the above signal model seems simple, there are numerous applications:

- Bayesian adaptive filtering [15], [16].
- Wireless channel estimation: multi-path parameter estimation as in [17], [18]. In this case, $\mathbf{x}_t$ = FIR filter response, and $\boldsymbol{\Gamma}$ represents e.g. the power delay profile.

III. VB-SBL

In Bayesian compressive sensing, a two-layer hierarchical prior is assumed for the $\mathbf{x}$ as in [4]. The hierarchical prior is such that it encourages the sparsity property of $\mathbf{x}_t$ or of the innovation sequences $\mathbf{v}_t$.

$$\begin{aligned} p(\mathbf{x}_t/\boldsymbol{\Gamma}) &= \prod_{i=1}^{M} p(x_{t,i}/\alpha_i) = \prod_{i=1}^{M} \mathcal{N}(0, \alpha_i^{-1}), \\ p(\mathbf{x}_t/\mathbf{x}_{t-1}, \mathbf{F}, \boldsymbol{\Gamma}) &= \prod_{i=1}^{M} p(x_{t,i}/x_{t-1,i}, \alpha_i, a_i) = \\ &\prod_{i=1}^{M} \mathcal{N}(a_i x_{t-1,i}, \frac{1}{\alpha_i}). \end{aligned} \tag{4}$$

Further a Gamma prior is considered over $\boldsymbol{\Gamma}$,

$$p(\boldsymbol{\Gamma}) = \prod_{i=1}^{M} p(\alpha_i/a, b) = \prod_{i=1}^{M} \Gamma^{-1}(a) b^a \alpha_i^{a-1} e^{-b\alpha_i}. \tag{5}$$

The inverse of noise variance γ is also assumed to have a Gamma prior,

$$p(\gamma) = \Gamma^{-1}(c) d^c \gamma_i^{c-1} e^{-d\gamma}. \tag{6}$$

Now the likelihood distribution can be written as,

$$p(\mathbf{y}_t/\mathbf{x}_t, \gamma) = (2\pi)^{-N} \gamma^N e^{\frac{-\gamma||\mathbf{y}_t - \mathbf{A}_t\mathbf{x}_t||^2}{2}}. \tag{7}$$

A. Variational Bayes

The computation of the posterior distribution of the parameters is usually intractable. In order to address this issue, in variational Bayesian framework, the posterior distribution $p(\mathbf{x}_t, \boldsymbol{\Gamma}, \gamma/\mathbf{y}_{1:t})$ is approximated by a variational distribution $q(\mathbf{x}_t, \boldsymbol{\Gamma}, \gamma)$ that has the factorized form:

$$q(\mathbf{x}_t, \boldsymbol{\Gamma}, \gamma) \ = \ q_\gamma(\gamma) \prod_{i=1}^{M} q_{x_{t,i}}(x_{t,i}) \prod_{i=1}^{M} q_{\alpha_i}(\alpha_i), \tag{8}$$

where $\mathbf{y}_{1:t}$ represents the observations till the time t $(\mathbf{y}_1, ..., \mathbf{y}_t)$, similarly we define $\mathbf{x}_{1:t}$. Variational Bayes compute the factors q by minimizing the Kullback-Leibler distance between the true posterior distribution $p(\mathbf{x}_t, \boldsymbol{\Gamma}, \gamma/\mathbf{y}_{1:t})$ and the $q(\mathbf{x}, \boldsymbol{\Gamma}, \gamma)$. From [10],

$$KLD_{VB} \ = \ KL\left(p(\mathbf{x}_t, \boldsymbol{\Gamma}, \gamma/\mathbf{y}_{1:t}) || q(\mathbf{x}_t, \boldsymbol{\Gamma}, \gamma)\right) \tag{9}$$

The KL divergence minimization is equivalent to maximizing the evidence lower bound (ELBO) [11]. To elaborate on this, we can write the marginal probability of the observed data as,

$$\begin{aligned} &\ln p(\mathbf{y}_t/\mathbf{y}_{1:t-1}) = L(q) + KLD_{VB}, \ \text{ where,} \\ &L(q) = \textstyle\int q(\mathbf{x}_t, \boldsymbol{\theta}) \ln \frac{p(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}/\mathbf{y}_{1:t-1})}{q(\boldsymbol{\theta})} d\mathbf{x}_t d\boldsymbol{\theta}, \\ &KLD_{VB} = -\textstyle\int q(\mathbf{x}_t, \boldsymbol{\theta}) \ln \frac{p(\mathbf{x}_t, \boldsymbol{\theta}/\mathbf{y}_{1:t})}{q(\mathbf{x}_t, \boldsymbol{\theta})} d\mathbf{x}_t d\boldsymbol{\theta}, \end{aligned} \tag{10}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\Gamma}, \gamma\}$ and θ_i represents each scalar in $\boldsymbol{\theta}$. Since $KLD_{VB} \geq 0$, it implies that $L(q)$ is a lower bound on $\ln p(\mathbf{y}_t/\mathbf{y}_{1:t-1})$. Moreover, $\ln p(\mathbf{y}_t/\mathbf{y}_{1:t-1})$ is independent of $q(\mathbf{x}_t, \boldsymbol{\theta})$ and therefore maximizing $L(q)$ is equivalent to minimizing KLD_{VB}. This is called as ELBO maximization and doing this in an alternating fashion for each variable in $\mathbf{x}_t, \boldsymbol{\theta}$ leads to,

$$\begin{aligned} &\ln(q_i(\theta_i)) = < \ln p(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}/\mathbf{y}_{1:t-1}) >_{\boldsymbol{\theta}_{\bar{i}}, \mathbf{x}_t} + c_i, \\ &\ln(q_i(x_{t,i})) = < \ln p(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}/\mathbf{y}_{1:t-1}) >_{\boldsymbol{\theta}, \mathbf{x}_{t,\bar{i}}} + c_i, \\ &p(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}/\mathbf{y}_{1:t-1}) = p(\mathbf{y}_t/\mathbf{x}_t, \gamma, \mathbf{y}_{1:t-1}) \\ &p(\mathbf{x}_t/\boldsymbol{\Lambda}, \mathbf{y}_{1:t-1}) p(\boldsymbol{\Lambda}) p(\gamma). \end{aligned} \tag{11}$$

ISBN 978-90-827970-1-5 © EURASIP 2018

Here $<>_{k\neq i}$ represents the expectation operator over the distributions q_k for all $k \neq i$. $\mathbf{x}_{t,\bar{i}}$ represents $\mathbf{x}_t$ without x_i and $\boldsymbol{\theta}_{\bar{i}}$ represents $\boldsymbol{\theta}$ without θ_i.

IV. SAVE Sparse Bayesian Learning and Kalman Filtering

In this section, we propose a Space Alternating Variational Estimation (SAVE) based alternating optimization between each elements of $\mathbf{x}_t$ and γ. For SAVE, not any particular structure of $\mathbf{A}_t$ is assumed, in contrast to AMP which performs poorly when $\mathbf{A}_t$ is not i.i.d or sub-Gaussian. The joint distribution w.r.t the observation of (2) can be written as,

$$p(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}/\mathbf{y}_{1:t-1}) = p(\mathbf{y}_t/\mathbf{x}_t, \boldsymbol{\theta}) p(\mathbf{x}_t, \boldsymbol{\theta}/\mathbf{y}_{1:t-1}), \quad (12)$$

where the predictive distribution $p(\mathbf{x}_t, \boldsymbol{\theta}/\mathbf{y}_{1:t-1})$ can be assumed to gaussian distribution with mean $\widehat{\mathbf{x}}_{t|t-1}$ and diagonal error covariance $\widehat{\mathbf{P}}_{t|t-1}$, $\mathcal{CN}(\widehat{\mathbf{x}}_{t|t-1}, \widehat{\mathbf{P}}_{t|t-1})$. Each diagonal entry of $\widehat{\mathbf{P}}_{t|t-1}$ is denoted as $\sigma^2_{t,k|t-1}$. $\widehat{\mathbf{x}}_{t|t-1}$ is denoted as the prediction (estimation of $\mathbf{x}_t$ from $\mathbf{y}_{1:t-1}$).

$$\begin{aligned}&\ln p(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}/\mathbf{y}_{1:t-1}) = N \ln \gamma - \gamma \left|\left|\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t\right|\right|^2 + \\ &-M \det(\widehat{\mathbf{P}}_{t|t-1}) - \left(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1}\right)^T \widehat{\mathbf{P}}^{-1}_{t|t-1} \left(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1}\right) \\ &+(c-1)\ln\gamma + c \ln d - d\gamma + \text{constants},\end{aligned} \quad (13)$$

In the following, $c_{x_{t,k}}, c'_{x_{t,k}}, c_{\alpha_k}$, c_{λ_k} and c_γ represents normalization constants for the respective pdfs.

A. Prediction Stage

In this stage, we compute the prediction about $\mathbf{x}_t$ at time $t-1$, $\widehat{x}_{t,k|t-1}$. In the prediction step of Kalman, like in [19] and based on the Chapman-Kolmogorov formula:

$$\begin{aligned}&p(\mathbf{x}_t, \boldsymbol{\Gamma}, \gamma | \mathbf{y}_{1:t-1}) = \\ &\int p(\mathbf{x}_t | \mathbf{x}_{t-1}, \boldsymbol{\Gamma}, \gamma, \mathbf{y}_{1:t-1}) p(\mathbf{x}_{t-1}, \boldsymbol{\Gamma}, \gamma | \mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1}\end{aligned} \quad (14)$$

Using variational approximation to the posterior $p(\mathbf{x}_{t-1}, \boldsymbol{\Gamma}, \gamma | \mathbf{y}_{1:t-1}) \sim q(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) q(\boldsymbol{\Gamma}, \gamma | \mathbf{y}_{1:t-1})$, the expression above results in

$$\begin{aligned}&p(\mathbf{x}_t, \boldsymbol{\Gamma}, \gamma | \mathbf{y}_{1:t-1}) \sim q(\boldsymbol{\Gamma}, \gamma | \mathbf{y}_{1:t-1}) \\ &\int p(\mathbf{x}_t | \mathbf{x}_{t-1}, \boldsymbol{\Gamma}, \mathbf{y}_{1:t-1}) q(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1} \\ &\sim q(\boldsymbol{\Gamma}, \gamma | \mathbf{y}_{1:t-1}) q(\mathbf{x}_t | \mathbf{y}_{1:t-1})\end{aligned} \quad (15)$$

Using VB we approximate $q(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ like the following,

$$\begin{aligned}&\ln q(\mathbf{x}_t | \mathbf{y}_{1:t-1}) \sim \mathrm{E}_{q(\boldsymbol{\Gamma}, \gamma | \mathbf{y}_{1:t-1})} \\ &\left(\ln\left(\int p(\mathbf{x}_t | \mathbf{x}_{t-1}, \boldsymbol{\Gamma}, \gamma, \mathbf{y}_{1:t-1}) q(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1}\right)\right)\end{aligned} \quad (16)$$

Defining $\widehat{f}_{k|t-1}$ as the mean of the approximate posterior for f_k and the variance as $\sigma^2_{a_{k|t-1}}$ at time instant $t-1$. From the time update equation of the standard Kalman filter,

$$\begin{aligned}&x_{t,k} = a_k \widehat{x}_{t-1,k|t-1} + v_k, \\ &\langle \tfrac{1}{|a_k|^2 \sigma^2_{t-1,k|t-1} + \frac{1}{\lambda_k}} (|x_{t,k}|^2 - x^H_{t,k} a_k \widehat{x}_{t-1,k|t-1} - \\ &a^H_k \widehat{x}_{t-1,k|t-1} x_{t,k} + |a_k|^2 |\widehat{x}_{t-1,k|t-1}|^2) \rangle \\ &= \tfrac{1}{\sigma^2_{t,k|t-1}} |x_{t,k} - \widehat{x}_{t,k|t-1}|^2\end{aligned} \quad (17)$$

To compute the term $\langle \frac{f_k}{\frac{1}{\lambda_k} + |f_k|^2 \sigma^2_{t-1,k|t-1}} \rangle$, we write $f_k = \widehat{f}_{k|t-1} + \widetilde{a}_{k|t-1}$, where $\widetilde{a}_{k|t-1}$ is a complex gaussian random variable with variance $\sigma^2_{f_k|t-1}$,

$$\begin{aligned}&\tfrac{f_k}{\frac{1}{\lambda_k} + |f_k|^2 \sigma^2_{t-1,k|t-1} + \frac{1}{\lambda_k}} = \\ &\tfrac{\widehat{f}_{k|t-1} + \widetilde{a}_{k|t-1}}{(|a_{k|t-1}|^2 + \widehat{f}_{k|t-1} \widetilde{a}^H_{k|t-1} + \widehat{f}^H_{k|t-1} \widetilde{a}_{k|t-1} + |\widetilde{a}_{k|t-1}|^2) \sigma^2_{t-1,k|t-1} + \frac{1}{\lambda_k}}, \\ &= \tfrac{\widehat{f}_{k|t-1} + \widetilde{a}_{k|t-1}}{(|a_{k|t-1}|^2 \sigma^2_{t-1,k|t-1} + \frac{1}{\lambda_k})(1 + \frac{(\widehat{f}_{k|t} \widetilde{a}^H_{k|t} + \widehat{f}^H_{k|t} \widetilde{a}_{k|t}) \sigma^2_{t-1,k|t-1}}{|a_{k|t-1}|^2 \sigma^2_{t-1,k|t-1} + \frac{1}{\lambda_k}}}, \\ &= \tfrac{\widehat{f}_{k|t-1}(1 - \frac{\sigma^2_{t-1,k|t-1} \sigma^2_{f_k|t-1}}{<|a_{k|t-1}|^2> \sigma^2_{t-1,k|t-1} + \frac{1}{<\lambda_k>}})}{<|a_{k|t-1}|^2> \sigma^2_{t-1,k|t-1} + \frac{1}{\lambda_k}}, \\ &= \tfrac{\widehat{f}_{k|t-1}}{<|a_{k|t-1}|^2> \sigma^2_{t-1,k|t-1} + \frac{1}{<\lambda_k>} + \sigma^2_{t-1,k|t-1} \sigma^2_{f_k|t-1}}}\end{aligned} \quad (18)$$

Finally we obtain,

$$\begin{aligned}&\sigma^2_{t,k|t-1} = (|\widehat{f}_{k|t-1}|^2 + \sigma^2_{a_{k|t-1}}) \sigma^2_{t-1,k|t-1} + \tfrac{1}{\widehat{\lambda}_{k|t-1}}, \\ &\widehat{x}_{t,k|t-1} = \widehat{f}_{k|t-1} \widehat{x}_{t-1,k|t-1}\end{aligned} \quad (19)$$

B. Measurement or Update Stage

Update of $q_{x_{t,k}}(x_{t,k})$: Using (11), $\ln q_{x_{t,k}}(x_{t,k})$ turns out to be quadratic in $x_{t,k}$ and thus can be represented as a Gaussian distribution as follows,

$$\begin{aligned}&\ln q_{x_{t,k}}(x_{t,k}) = - <\gamma> \Big\{ (\mathbf{y}_t - \mathbf{A}_{t,\bar{k}} < \mathbf{x}_{t,\bar{k}} >)^H \mathbf{A}_{t,k} x_{t,k} - \\ &x^H_{t,k} \mathbf{A}^H_{t,k} (\mathbf{y}_t - \mathbf{A}_{t,\bar{k}} < \mathbf{x}_{t,\bar{k}} >) + ||\mathbf{A}_{t,k}||^2 |x_{t,k}|^2 \Big\} - \\ &\tfrac{1}{2\sigma^2{}_{t,k|t-1}} \left(|x_{t,k}|^2 - x^H_{t,k} \widehat{x}_{t,k|t-1} - x_{t,k} \widehat{x}^H_{t,k|t-1}\right) + c_{x_{t,k}} = \\ &-\tfrac{1}{\sigma^2_{t,k|t}} \left|x_{t,k} - \widehat{x}_{t,k|t}\right|^2 + c'_{x_{t,k}}.\end{aligned} \quad (20)$$

Note that we split $\mathbf{A}_t \mathbf{x}_t$ as, $\mathbf{A}_t \mathbf{x}_t = \mathbf{A}_{t,k} x_{t,k} + \mathbf{A}_{t,\bar{k}} \mathbf{x}_{t,\bar{k}}$, where $\mathbf{A}_{t,k}$ represents the k^{th} column of $\mathbf{A}_t$, $\mathbf{A}_{t,\bar{k}}$ represents the matrix with k^{th} column of $\mathbf{A}_t$ removed. Clearly, the mean and the variance of the resulting Gaussian distribution becomes,

$$\begin{aligned}&\sigma^2_{t,k|t} = \tfrac{1}{<\gamma> ||\mathbf{A}_{t,k}||^2 + \frac{1}{\sigma^2_{t,k|t-1}}}, \\ &\widehat{x}_{t,k|t} = \\ &\sigma^2_{t,k|t} \left(\mathbf{A}^H_{t,k} \left(\mathbf{y}_t - \mathbf{A}_{t,\bar{k}} < \mathbf{x}_{t,\bar{k}} >\right) <\gamma> + \tfrac{\widehat{x}_{t,k|t-1}}{\sigma^2_{t,k|t-1}}\right),\end{aligned} \quad (21)$$

where $\widehat{x}_{t,k|t}$ represents the point estimate of $x_{t,k}$. One remark is that forcing a Gaussian posterior q with diagonal covariance matrix on the original Kalman measurement equations gives the same result as SAVE.

Update of $q_\gamma(\gamma)$: Similarly, the Gamma distribution from the variational Bayesian approximation for the $q_\gamma(\gamma)$ can be written as,

$$\begin{aligned}&\ln q_\gamma(\gamma) = \\ &(c - 1 + N) \ln \gamma - \gamma \left(< ||\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t||^2 > + d\right) + c_\gamma, \\ &q_\gamma(\gamma) \propto \gamma^{c+N-1} e^{-\gamma \left(<||\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t||^2> + d\right)}.\end{aligned} \quad (22)$$

ISBN 978-90-827970-1-5 © EURASIP 2018

The mean of the Gamma distribution for γ is given by,

$$\begin{aligned}
&< \gamma >= \widehat{\gamma}_t = \tfrac{c+\frac{N}{2}}{(\xi_t+d)},\\
&\xi_t = \beta\xi_{t-1} + (1-\beta) < ||\mathbf{y}_t - \mathbf{A}_t\mathbf{x}_t||^2 >, \text{ where,}\\
&< ||\mathbf{y}_t - \mathbf{A}_t\mathbf{x}_t||^2 >=\\
&||\mathbf{y}_t||^2 - 2\Re(\mathbf{y}_t^H\mathbf{A}_t\widehat{\mathbf{x}}_{t|t}) + \text{tr}\left(\mathbf{A}_t^H\mathbf{A}_t(\widehat{\mathbf{x}}_{t|t}\widehat{\mathbf{x}}_{t|t}^H + \boldsymbol{\Sigma}_{t|t})\right),\\
&\boldsymbol{\Sigma}_{\mathbf{t|t}} = diag(\sigma_{t,1|t}^2, ..., \sigma_{t,M|t}^2),\ \widehat{\mathbf{x}}_{t|t} = [\widehat{x}_{t,1|t}, ..., \widehat{x}_{t,M|t}]^T,
\end{aligned} \tag{23}$$

where we introduced temporal averaging also and β denotes the weighting coefficients which are less than one.

C. Fixed Lag Smoothing

For fixed lag smoothing with delay $\Delta > 0$, we rewrite the state space model as follows,

$$\begin{aligned}
&\mathbf{y}_t = \mathbf{A}_t\mathbf{F}^{\Delta}\mathbf{x}_{t-\Delta} + \underbrace{\sum_{i=0}^{\Delta-1}\mathbf{A}_t\mathbf{F}^i\mathbf{v}_{t-i} + \mathbf{w}_t}_{\widetilde{\mathbf{w}}_t},\\
&p(\mathbf{y}_t, \mathbf{x}_{t-\Delta}, \boldsymbol{\theta}/\mathbf{y}_{1:t-1}) = p(\mathbf{y}_t/\mathbf{x}_{t-\Delta}, \boldsymbol{\theta})p(\mathbf{x}_{t-\Delta}, \boldsymbol{\theta}/\mathbf{y}_{1:t-1}),
\end{aligned} \tag{24}$$

where $\widetilde{\mathbf{w}}_t \sim \mathcal{CN}(\mathbf{0}, \widetilde{\mathbf{R}}_t)$, $\widetilde{\mathbf{R}}_t = \mathbf{A}_t(\mathbf{I} - |\mathbf{F}|^{2\Delta})\boldsymbol{\Gamma})\mathbf{A}_t^H + \frac{1}{\gamma}\mathbf{I}$. The posterior distribution $p(\mathbf{x}_{t-\Delta}, \boldsymbol{\theta}/\mathbf{y}_{1:t-1})$ is approximated using variational approximation as $q(\mathbf{x}_{t-\Delta}, \boldsymbol{\theta}/\mathbf{y}_{1:t-1})$ with mean and covariance as $\widehat{\mathbf{x}}_{t-\Delta|t-1}$ and $\boldsymbol{\Sigma}_{t-\Delta|t-1}$.

$$\begin{aligned}
&\ln p(\mathbf{y}_t, \mathbf{x}_{t-\Delta}, \boldsymbol{\theta}/\mathbf{y}_{1:t-1}) = -\tfrac{1}{2}\ln\det\widetilde{\mathbf{R}}_t -\\
&(\mathbf{y}_t - \mathbf{A}_t\mathbf{F}^{\Delta}\mathbf{x}_{t-\Delta})^H\widetilde{\mathbf{R}}_t^{-1}(\mathbf{y}_t - \mathbf{A}_t\mathbf{F}^{\Delta}\mathbf{x}_{t-\Delta})\\
&-\tfrac{1}{2}\det(\boldsymbol{\Sigma}_{t-\Delta|t-1}) -\\
&\left(\mathbf{x}_{t-\Delta} - \widehat{\mathbf{x}}_{t-\Delta|t-1}\right)^H\boldsymbol{\Sigma}_{t-\Delta|t-1}^{-1}\left(\mathbf{x}_{t-\Delta} - \widehat{\mathbf{x}}_{t-\Delta|t-1}\right) + c_{x-\Delta}.
\end{aligned} \tag{25}$$

Here $|\mathbf{F}|$ represents the matrix obtained by the amplitudes of the entries in $\mathbf{F}$ and $c_{x-\Delta}$ represents normalization constants. **Update of $\mathbf{x}_{t-\Delta}$:**Using (11), $\ln q_{\mathbf{x}_{t-\Delta}}(\mathbf{x}_{t-\Delta}/\mathbf{y}_{1:t})$ turns out to be quadratic in $\mathbf{x}_{t-\Delta}$ and thus can be represented as a Gaussian distribution with mean and covariance as $\widehat{\mathbf{x}}_{t-\Delta|t}$ and $\boldsymbol{\Sigma}_{t-\Delta|t}$ respectively,

$$\begin{aligned}
&\boldsymbol{\Sigma}_{t-\Delta|t} = (\mathbf{F}^{\Delta\,H}\mathbf{A}_t^H < \widetilde{\mathbf{R}}_t^{-1} > \mathbf{A}_t\mathbf{F}^{\Delta} + \boldsymbol{\Sigma}_{t-\Delta|t-1}^{-1})^{-1},\\
&\widehat{\mathbf{x}}_{t-\Delta|t} =\\
&\boldsymbol{\Sigma}_{t-\Delta|t}(\boldsymbol{\Sigma}_{t-\Delta|t-1}^{-1}\widehat{\mathbf{x}}_{t-\Delta|t-1} + \mathbf{F}^{\Delta\,H}\mathbf{A}_t^H < \widetilde{\mathbf{R}}_t^{-1} > \mathbf{y}_t).
\end{aligned} \tag{26}$$

D. VB for AR(1) Parameters

In this section VB is used to learn the unknown correlation coefficient f_k and λ_k. The goal is to maximize the posterior $p(f_k/\mathbf{x}_t, \mathbf{y}_{1:t})$ and $p(\lambda_k/\mathbf{x}_t, \mathbf{y}_{1:t})$. We denote $\widehat{f}_{k|t-1}$ as the mean or the point estimate of f_k at time instant $t-1$. The update of correlation coefficient at t can be written as, From the state space model, $x_{t,k} = f_k x_{t-1,k} + \frac{1}{\sqrt{\lambda_k}}v_{t,k}$,

$$\begin{aligned}
&\ln p(\mathbf{x}_t, \mathbf{x}_{t-1}\boldsymbol{\Gamma}, f_k, \mathbf{y}_t/\mathbf{y}_{1:t-1}) = \ln\lambda_k -\\
&\lambda_k\left|x_{t,k} - f_k x_{t-1,k}\right|^2 + ((a-1)\ln\lambda_k + a\ln b - b\lambda_k)\\
&+\text{“constants”},
\end{aligned} \tag{27}$$

where “constants” denote terms independent of λ_k and f_k.

Update of $q_{\lambda_k}(\lambda_k)$: Using variational approximation $\ln q(\lambda_k|\mathbf{y}_{1:t}) \sim \mathrm{E}_{q(\mathbf{x}_t, \mathbf{x}_{t-1}, \gamma, f_k/\mathbf{y}_{1:t})}\ln p(\mathbf{x}_t, \boldsymbol{\Gamma}, \gamma, f_k\mathbf{y}_{1:t})$,

$$\begin{aligned}
&\ln\lambda_k - \lambda_k(< |x_{t,k} - f_k x_{t-1,k}|^2 > +b) + (a-1)\ln\lambda_k\\
&+c_{\lambda_k},\\
&q_{\lambda_k}(\lambda_k) \propto \lambda_k^a e^{-\lambda_k\left(<|x_{t,k}-f_k x_{t-1,k}|^2>+b\right)}
\end{aligned} \tag{28}$$

The mean and variance of the resulting gamma distribution can be written as,

$$\begin{aligned}
&< \lambda_k >= \tfrac{(a+1)}{(<|x_{t,k}-f_k x_{t-1,k}|^2>+b)},\\
&< \beta_k >= \tfrac{(a+1)}{(<|x_{t,k}-f_k x_{t-1,k}|^2>+b)^2}
\end{aligned} \tag{29}$$

Update of $q_{f_k}(f_k)$: Using variational approximation $\ln q(f_k|\mathbf{y}_{1:t}) \sim \mathrm{E}_{q(\mathbf{x}_t, \mathbf{x}_{t-1}, \gamma/\mathbf{y}_{1:t})}\ln p(\mathbf{x}_t, \mathbf{x}_{t-1}, \boldsymbol{\Gamma}, \gamma, \mathbf{y}_{1:t})$,

$$q_{f_k}(f_k) \propto e^{-<|x_{t-1,k}|^2>|f_k - <x_{t,k}><x_{t-1,k}^H>|} \tag{30}$$

Finally we write the mean and variance as,

$$\begin{aligned}
&\sigma_{f_k|t}^2 = \tfrac{1}{\widehat{x}_{t-1,k|t}^2 + \sigma_{t-1,k|t}^2},\\
&\widehat{f}_{k|t} = \sigma_{f_k|t}^2\widehat{x}_{t-1,k|t}^H\widehat{x}_{t,k|t} = \tfrac{\widehat{x}_{t-1,k|t}^H\widehat{x}_{t,k|t}}{\widehat{x}_{t-1,k|t}^2 + \sigma_{t-1,k|t}^2}.
\end{aligned} \tag{31}$$

Algorithm 1: The Adaptive SAVE-KF Algorithm

Update Stage

$$\begin{aligned}
&\sigma_{t,k|t}^2 = \sigma_{t,k|t-1}^2(\sigma_{t,k|t-1}^2\widehat{\gamma}_{t-1}\,||\mathbf{A}_{t,k}|\,|^2 + 1)^{-1},\\
&\mathbf{K}_{t,k} = \sigma_{t,k|t}^2\mathbf{A}_{t,k}^H\widehat{\gamma}_{t-1},\\
&\widehat{x}_{t,k|t} = \tfrac{\sigma_{t,k|t}^2}{\sigma_{t,k|t-1}^2}\widehat{x}_{t,k|t-1} + \mathbf{K}_{t,k}\left(\mathbf{y}_t - \mathbf{A}_{t,\overline{k}}\widehat{\mathbf{x}}_{t,\overline{k}|t}\right),\\
&\widehat{\gamma}_t = \tfrac{c+\frac{N}{2}}{(\xi_t+d)},\\
&\eta_{k|t} = \Re\{\widehat{f}_{k|t-1}^H\widehat{x}_{t,k|t}\widehat{x}_{t-1,k|t-1}\},\\
&\widehat{\lambda}_{k|t} =\\
&\tfrac{a+1}{(|\widehat{x}_{t,k|t}|^2 + \sigma_{t,k|t}^2 + |\widehat{f}_{k|t-1}|^2(|\widehat{x}_{t-1,k|t-1}|^2 + |\widehat{\sigma}_{t-1,k|t-1}^2|) - 2\eta_{k|t} + b)}.
\end{aligned}$$

Prediction Stage

$$\begin{aligned}
&\sigma_{t+1,k|t}^2 = (|\widehat{f}_{k|t}|^2 + \sigma_{a_{k|t}}^2)\sigma_{t,k|t}^2 + \tfrac{1}{\widehat{\lambda}_{k|t}},\\
&\widehat{x}_{t+1,k|t} = \widehat{f}_{k|t}\widehat{x}_{t,k|t},
\end{aligned}$$

Estimation of AR(1) Parameters

$$\begin{aligned}
&\sigma_{f_k|t}^2 = \tfrac{1}{\widehat{x}_{t-1,k|t}^2 + \sigma_{t-1,k|t}^2},\\
&\widehat{f}_{k|t} = \sigma_{f_k|t}^2\widehat{x}_{t-1,k|t}^H\widehat{x}_{t,k|t} = \tfrac{\widehat{x}_{t-1,k|t}^H\widehat{x}_{t,k|t}}{\widehat{x}_{t-1,k|t}^2 + \sigma_{t-1,k|t}^2}.
\end{aligned} \tag{32}$$

E. Computational Complexity

For our proposed SAVE, it is evident that we don't need any matrix inversions compared to [4]. Our computational complexity is similar to [12]. Update of all the variable $\mathbf{x}_t, \boldsymbol{\Gamma}, \gamma$ involves simple addition and multiplication operations. We introduce the following variables, $\mathbf{q} = \mathbf{y}_t^H\mathbf{A}_t$ and $\mathbf{B} = \mathbf{A}_t^H\mathbf{A}_t$. $\mathbf{q}, \mathbf{B}$ and $||\mathbf{y}_t|\,|^2$ can be precomputed, so only computed once.

V. Simulation Results

In this section we present the simulation results to validate the performance of our SAVE-KF SBL algorithm (Algorithm 1) compared to the state of the art solutions. We consider the special case of uncorrelated measurements $a_i = 0, \forall i$, thus

ISBN 978-90-827970-1-5 © EURASIP 2018

reducing it to the single measurement vector case. We compare our algorithm with the Fast Inverse-Free SBL (Fast IF SBL) in [12], the G-AMP based SBL in [20] and the fast version of SBL (FV SBL) in [13]. For the simulations, we have fixed $M = 200$ and $K = 30$. All the elements of $\mathbf{A}_t$ and $\mathbf{x}_t$ are generated i.i.d from a normal distribution, $\mathcal{N}(0, 1)$. SNR is fixed to be $20\,dB$ in the simulation.

A. MSE Performance

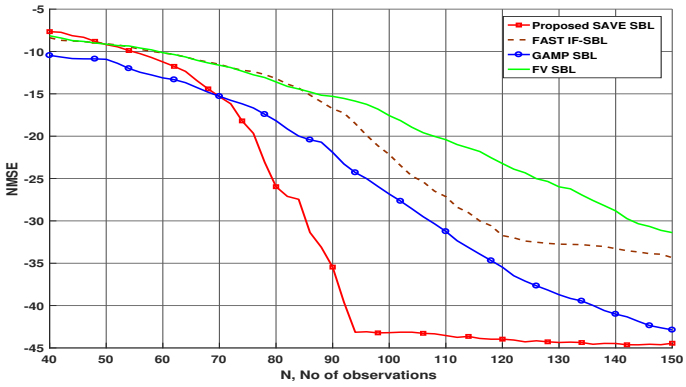

Fig. 1. NMSE vs the number of observations.

From Figure 1, it is evident that proposed SAVE algorithm performs better than the state of the art solutions in terms of the Normalized Mean Square Error (NMSE). NMSE is defined as $NMSE = \frac{1}{M}\,||\widehat{\mathbf{x}} - \mathbf{x}|\,|^2$, $\widehat{\mathbf{x}}$ represents the estimated value, $NMSE_{dB} = 10\log 10(NMSE)$.

B. Complexity

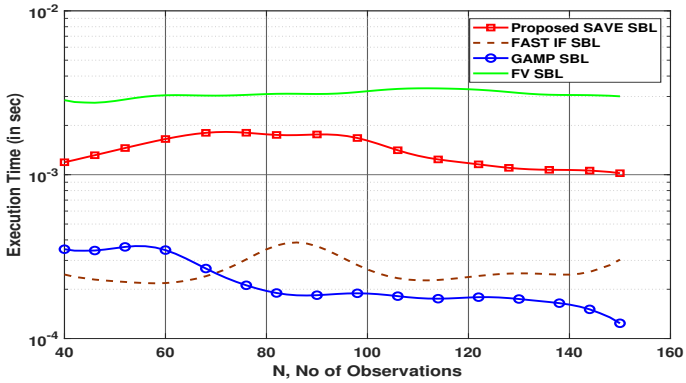

Fig. 2. Number of iterations vs the number of observations.

Since the proposed SAVE have similar computational requirements compared to [12], [20] , we plot the execution time in matlab for all the algorithms. It is clear from Figure 2 that proposed SAVE approach has lower exection time and thus faster convergence compared to the existing fast SBL algorithm. However, GAMP based [20] and IF-SBL [12] has a better convergence rate compared to ours but with a degradation in NMSE performance.

VI. Conclusion

We presented a fast SBL algorithm called SAVE-KF, which uses the variational inference techniques to approximate the posteriors of the data and parameters and track a time varying sparse signal. SAVE-KF helps to circumvent the matrix inversion operation required in conventional SBL using EM algorithm. We showed that the proposed algorithm has a faster execution time or convergence rate and better performance in terms of NMSE than even the state of the art fast SBL solutions. SAVE-KF algorithm exploit the underlying sparsity in the signal compared to the classical Kalman filtering based methods.

References

[1] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit ," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, December 2007.

[2] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit ," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[3] D. Wipf and S. Nagarajan, "Iterative reweighted l_1 and l_2 methods for finding sparse solutions ," *IEEE J. Sel. Topics Sig. Process.*, vol. 4, no. 2, pp. 317–329, April 2010.

[4] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[5] D. P. Wipf and B. D. Rao, "Sparse Bayesian Learning for Basis Selection ," *IEEE Trans. on Sig. Process.*, vol. 52, no. 8, pp. 2153–2164, August 2004.

[6] Z. Zhang and B. D. Rao, "Sparse Signal Recovery with Temporally Correlated Source Vectors Using Sparse Bayesian Learning ," *IEEE J. of Sel. Topics in Sig. Process.*, vol. 5, no. 5, pp. 912 – 926, September 2011.

[7] R. Prasad, C. Murthy, and B. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse Bayesian learning ," *IEEE Trans. on Sig. Process.*, vol. 62, no. 14, July 2014.

[8] R. Giri and B. D. Rao, "Type I and type II bayesian methods for sparse signal recovery using scale mixtures," *IEEE Trans. on Sig Process.*, vol. 64, no. 13, pp. 3418–3428, 2018.

[9] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *AISTATS*, January 2003.

[10] M. J. Beal, "Variational algorithms for approximate bayesian inference," in *Thesis, Univeristy of Cambridge, UK*, May 2003.

[11] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Sig. Process. Mag.*, vol. 29, no. 6, pp. 131–146, November 2008.

[12] H. Duan, L. Yang, and H. Li, "Fast inverse-free sparse bayesian learning via relaxed evidence lower bound maximization," *IEEE Sig. Process. Letters*, vol. 24, no. 6, June 2017.

[13] D. Shutin, T. Buchgraber, S. R. Kulkarni, and H. V. Poor, "Fast variational sparse bayesian learningwith automatic relevance determination for superimposed signals," *IEEE Trans. on Sig. Process*, vol. 59, no. 12, December 2011.

[14] C. K. Thomas and D. Slock, "SAVE - space alternating variational estimation for sparse Bayesian learning," in *Data Science Workshop*, 2018.

[15] T. Sadiki and D. T. Slock, "Bayesian adaptive filtering: principles and practical approaches," in *EUSIPCO*, 2004.

[16] J. B. S. Ciochina, C. Paleologu, "A family of optimized LMS-based algorithms for system identification," in *Proc. EUSIPCO*, 2016, pp. 1803–1807.

[17] C. K. Thomas and D. Slock, "Variational bayesian learning for channel estimation and transceiver determination," in *Information Theory and Applications Workshop*, San Diego, USA, February 2018.

[18] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. I. Pedersen, "Channel Parameter Estimation in Mobile Radio Environments Using the SAGE Algorithm ," *IEEE J. on Sel. Areas in Commun.*, vol. 17, no. 3, pp. 434–450, March 1999.

[19] S. Sarkka and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational bayesian approximations ," *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 596–600, March 2009.

[20] M. Al-Shoukairi, P. Schniter, and B. D. Rao, "GAMP-based low complexity sparse bayesian learning algorithm," *IEEE Trans. on Sig. Process.*, vol. 66, no. 2, January 2018.

ISBN 978-90-827970-1-5 © EURASIP 2018